Definition 5.1. *A **vector** in $\mathbb{R}^n$ is an n-tuple $\vec{v} = (v_1, v_2, \ldots, v_n)$. The **zero vector** $\vec{0} = (0, 0, \ldots, 0)$. Given two vectors $\vec{v}$ and $\vec{w}$ in $\mathbb{R}^n$, the sum $\vec{v} + \vec{w}$ is the vector $(v_1 + w_1, v_2 + w_2, \ldots, v_n + w_n)$. If λ is a scalar, the **scalar product** $\lambda\vec{v}$ is the vector $(\lambda v_1, \lambda v_2, \ldots, \lambda v_n)$.*

The sum and scalar product of vectors in $\mathbb{R}^n$ obey the same rules as the sum and scalar product in $\mathbb{R}^2$ and $\mathbb{R}^3$.

Definition 5.2. *Let $\vec{v}$ and $\vec{w} \in \mathbb{R}^n$. The **dot product** is the scalar*

$$\vec{v} \cdot \vec{w} = v_1 w_1 + v_2 w_2 + \ldots v_n w_n.$$

*The **norm** (or **length**) of $\vec{v}$ is the scalar*

$$\|\vec{v}\| = \sqrt{\vec{v} \cdot \vec{v}}.$$

The scalar product obeys the usual rules.

Example 5.3. *Suppose that $\vec{v} = (1, 2, 3, 4)$ and $\vec{w} = (2, -1, 1, -1)$ and $\lambda = -2$. Then*

$$\vec{v} + \vec{w} = (3, 1, 4, 3) \qquad \text{and} \qquad \lambda\vec{w} = (-4, 2, -2, 2).$$

We have

$$\vec{v} \cdot \vec{w} = 2 - 2 + 3 - 4 = -1.$$

The standard basis of $\mathbb{R}^n$ is the set of vectors,

$$e_1 = (1, 0, \ldots, 0), \quad e_2 = (0, 1, \ldots, 0), \quad e_1 = (0, 0, 1, \ldots, 0), \quad \ldots e_n = (0, 0, \ldots, 1).$$

Note that if $\vec{v} = (v_1, v_2, \ldots, v_n)$, then

$$\vec{v} = v_1 e_1 + v_2 e_2 + \cdots + v_n e_n.$$

Let's adopt the (somewhat ad hoc) convention that $\vec{v}$ and $\vec{w}$ are parallel if and only if either $\vec{v}$ is a scalar multiple of $\vec{w}$, or vice-versa. Note that if both $\vec{v}$ and $\vec{w}$ are non-zero vectors, then $\vec{v}$ is a scalar multiple of $\vec{w}$ if and only if $\vec{w}$ is a scalar multiple of $\vec{v}$.

Theorem 5.4 (Cauchy-Schwarz-Bunjakowski)**.** *If $\vec{v}$ and $\vec{w}$ are two vectors in $\mathbb{R}^n$, then*

$$|\vec{v} \cdot \vec{w}| \leq \|v\|\|w\|,$$

with equality if and only if $\vec{v}$ is parallel to $\vec{w}$.

Proof. If either $\vec{v}$ or $\vec{w}$ is the zero vector, then there is nothing to prove. So we may assume that neither vector is the zero vector.

Let $\vec{u} = x\vec{v} + \vec{w}$, where x is a scalar. Then

$$0 \leq \vec{u} \cdot \vec{u} = (\vec{v} \cdot \vec{v})x^2 + 2(\vec{v} \cdot \vec{w})x + \vec{w} \cdot \vec{w} = ax^2 + bx + c.$$

So the quadratic function $f(x) = ax^2 + bx + c$ has at most one root. It follows that the discriminant is less than or equal to zero, with equality if and only if $f(x)$ has a root. So

$$4(\vec{v} \cdot \vec{w})^2 - 4\|\vec{v}\|^2\|\vec{w}\|^2 = b^2 - 4ac \leq 0.$$

Rearranging, gives

$$(\vec{v} \cdot \vec{w})^2 \leq \|\vec{v}\|^2\|\vec{w}\|^2.$$

Taking square roots, gives

$$|\vec{v} \cdot \vec{w}| \leq \|v\|\|w\|.$$

Now if we have equality here, then the discriminant must be equal to zero, in which case we may find a scalar λ such that the vector $\lambda\vec{v} + \vec{w}$ has zero length. But the only vector of length zero is the zero vector, so that $\lambda\vec{v}+\vec{w} = \vec{0}$. In other words, $\vec{w} = -\lambda\vec{v}$ and $\vec{v}$ and $\vec{w}$ are parallel. □

Definition 5.5. *If $\vec{v}$ and $\vec{w} \in \mathbb{R}^n$ are non-zero vectors, then the angle between them is the unique angle $0 \leq \theta \leq \pi$ such that*

$$\cos\theta = \frac{\vec{v} \cdot \vec{w}}{\|\vec{v}\|\|\vec{w}\|}.$$

Note that the fraction is between -1 and 1, by (5.4), so this does makes sense, and we showed in (5.4) that the angle is 0 or π if and only if $\vec{v}$ and $\vec{w}$ are parallel.

Definition 5.6. *If $A = (a_{ij})$ and $B = (b_{ij})$ are two $m \times n$ matrices, then the **sum** $A + B$ is the $m \times n$ matrix $(a_{ij} + b_{ij})$. If λ is a scalar, then the **scalar multiple** λA is the $m \times n$ matrix (λa_{ij}).*

Example 5.7. *If*

$$A = \begin{pmatrix} 1 & -1 \\ 3 & -4 \end{pmatrix}, \qquad \text{and} \qquad B = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix},$$

then

$$A + B = \begin{pmatrix} 2 & 0 \\ 5 & -5 \end{pmatrix},$$

and

$$3A = \begin{pmatrix} 3 & -3 \\ 9 & -12 \end{pmatrix}.$$

Note that if we *flattened* A and B to $(1, -1, 3, -4)$ and $(2, 0, 5, -5)$ then the sum corresponds to the usual vector sum $(3, -1, 8, -9)$. Ditto for scalar multiplication.

Definition 5.8. *Suppose that $A = (a_{ij})$ is an $m \times n$ matrix and $B = (b_{ij})$ is an $n \times p$ matrix. The* ***product*** $C = AB = (c_{ij})$ *is the $m \times p$ matrix where*

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + a_{i3}b_{3j} + \cdots + a_{in}b_{nj} = \sum_{l=1}^{n} a_{il}b_{lj}.$$

In other words, the entry in the ith row and jth column of C is the dot product of the ith row of A and the jth column of B. This only makes sense because the ith row and the jth column are both vectors in $\mathbb{R}^n$.

Example 5.9. *Let*

$$A = \begin{pmatrix} 1 & -2 & 1 \\ 1 & -1 & 5 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 2 & 1 \\ 1 & -4 \\ -1 & 1 \end{pmatrix}.$$

Then $C = AB$ has shape 2×2, and in fact

$$C = AB = \begin{pmatrix} -1 & 10 \\ -4 & 10 \end{pmatrix}.$$

Theorem 5.10. *Let A, B and C be three matrices, and let λ and μ be scalars.*

(1) *If A, B and C have the same shape, then $(A + B) + C = A + (B + C)$.*
(2) *If A and B have the same shape, then $A + B = B + A$.*
(3) *If A and B have the same shape, then $\lambda(A + B) = \lambda A + \lambda B$.*
(4) *If Z is the zero matrix with the same shape as A, then $Z + A = A + Z$.*
(5) $\lambda(\mu A) = (\lambda\mu)A$.
(6) $(\lambda + \mu)A = \lambda A + \mu A$.
(7) *If I_n is the matrix with ones on the diagonal and zeroes everywhere else and A has shape $m \times n$, then $AI_n = A$ and $I_m A = A$.*
(8) *If A has shape $m \times n$ and B has shape $n \times p$ and C has shape $p \times q$, then $A(BC) = (AB)C$.*
(9) *If A has shape $m \times n$ and B and C have the same shape $n \times p$, then $A(B + C) = AB + AC$.*
(10) *If A and B have the same shape $m \times n$ and C has shape $n \times p$, then $(A + B)C = AC + BC$.*

Example 5.11. *Note however that $AB \neq BA$ in general. For example if A has shape 1×3 and B has shape 3×2, then it makes sense to multiply A and B but it does not make sense to multiply B and A.*

In fact even if it makes sense to multiply A and B and B and A, the two products might not even have the same shape. For example, if

$$A = \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix},$$

and

$$B = \begin{pmatrix} 2 & -1 & 3 \end{pmatrix},$$

then AB has shape 3×3,

$$AB = \begin{pmatrix} 2 & -1 & 3 \\ 4 & -2 & 6 \\ -2 & 1 & -3 \end{pmatrix},$$

but BA has shape 1×1,

$$BA = (2 - 2 - 3) = (-3).$$

But even both products AB and BA make sense, and they have the same shape, the products still don't have to be equal. Suppose

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \qquad \textit{and} \qquad B = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Then AB and BA are both 2×2 matrices. But

$$AB = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix}, \qquad \textit{and} \qquad BA = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix}.$$

One can also define determinants for $n \times n$ matrices. It is probably easiest to explain the general rule using an example:

$$\begin{vmatrix} 1 & 0 & 0 & 2 \\ 2 & 0 & 1 & -1 \\ 1 & -2 & 1 & 1 \\ 0 & 1 & 0 & 1 \end{vmatrix} = \begin{vmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \\ 1 & 0 & 1 \end{vmatrix} - 2 \begin{vmatrix} 2 & 0 & 1 \\ 1 & -2 & 1 \\ 0 & 1 & 0 \end{vmatrix}.$$

Notice that we as expand about the top row, the sign alternates $+-+-$, so that the last term comes with a minus sign.

Finally, we try to explain the real meaning of a matrix. Let

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}.$$

Given A, we can construct a function

$$f \colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2,$$

by the rule

$$f(\vec{v}) = A\vec{v}.$$

If $\vec{v} = (x, y)$, then

$$A\vec{v} = (x + y, y).$$

Here I cheat a little, and write row vectors instead of column vectors. Geometrically this is called a *shear*; it leaves the y-axis alone but one goes further along the x-axis according to the value of y. If

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$

the resulting function sends (x, y) to $(ax + by, cx + dy)$. In fact the functions one gets this way are always linear. If

$$A = \begin{pmatrix} 2 & 0 \\ 0 & -1 \end{pmatrix},$$

then $f(x, y) = (2x - y)$, and this has the result of scaling by a factor of 2 in the x-direction and reflects in the y-direction.

In general if A is an $m \times n$ matrix, we get a function

$$f \colon \mathbb{R}^n \longrightarrow \mathbb{R}^m,$$

using the same rule, $f(\vec{v}) = A\vec{v}$. If B is an $n \times p$ matrix, then we get a function

$$g \colon \mathbb{R}^p \longrightarrow \mathbb{R}^n,$$

by the rule $g(\vec{w}) = B\vec{w}$. Note that we can compose the functions f and g, to get a function

$$f \circ g \colon \mathbb{R}^p \longrightarrow \mathbb{R}^m.$$

First we apply g to $\vec{w}$ to get a vector $\vec{v}$ in $\mathbb{R}^n$ and then we apply f to $\vec{v}$ to get a vector in $\mathbb{R}^m$. The composite function $f \circ g$ is given by the rule $(f \circ g)(\vec{w}) = (AB)\vec{w}$. In other words, matrix multiplication is chosen so that it represents composition of functions.

As soon as one realises this, many aspects of matrix multiplication become far less mysterious. For example, composition of functions is not commutative, for example

$$\sin 2x \neq 2 \sin x,$$

and this is why $AB \neq BA$ in general. Note that it is not hard to check that composition of functions is associative,

$$f \circ (g \circ h) = (f \circ g) \circ h.$$

This is the easiest way to check that matrix multiplication is associative, that is, (8) of (5.10).

Functions given by matrices are obviously very special. Note that if $f(\vec{v}) = A\vec{v}$, then

$$f(\vec{v}+\vec{w}) = A(\vec{v}+\vec{w}) = A\vec{v} + A\vec{w} = f(\vec{v}) + f(\vec{w}),$$

and

$$f(\lambda\vec{v}) = A(\lambda\vec{v}) = \lambda(A\vec{v}) = \lambda f(\vec{v}).$$

Any function which respects both addition of vectors and scalar multiplication is called **linear** and it is precisely the linear functions which are given by matrices. In fact if $e_1, e_2, \dots, e_n$ and $f_1, f_2, \dots, f_m$ are standard bases for $\mathbb{R}^n$ and $\mathbb{R}^m$, and f is linear, then

$$f(e_j) = \sum a_{ij} f_i,$$

for some scalars a_{ij}, since $f(e_j)$ is a vector in $\mathbb{R}^m$ and any vector in $\mathbb{R}^m$ is a linear combinaton of the standard basis vectors $f_1, f_2, \dots, f_m$. If we put $A = (a_{ij})$ then one can check that f is the function

$$f(\vec{v}) = A\vec{v}.$$